

## ИСПОЛЬЗОВАНИЕ ОНТОЛОГИЧЕСКИХ ЗНАНИЙ В РЕКОМЕНДУЮЩИХ СИСТЕМАХ

Проанализированы основные функции рекомендующих систем и критерии их классификации. Предложены методы выработки рекомендаций на основе онтологической модели знаний о пользователях и ресурсах и ее использование в семантической поисковой системе МАИПС.

### Введение

Постоянное увеличение объема контента, доступного через World Wide Web, его гетерогенность, слабая структурированность и нечеткость классификации усложняют его эффективное использование. У большинства пользователей возникает необходимость в фильтрации доступной через Web информации таким образом, чтобы найти то, что соответствует реальным информационным потребностям пользователя. Для этого пользователю необходимо как-то задать предметную область своих интересов и классифицировать искомый объект (таким объектом может быть, к примеру, информационный ресурс, сервис или товар, продаваемый через Интернет).

Именно этим занимаются различные рекомендующие системы, ориентированные на различные типы ресурсов и предметные области.

К сожалению, часто людям сложно четко сформулировать, чем они интересуются и что именно им нужно (причем проблема не сводится только к используемой терминологии), однако им значительно проще оценить, подходит ли им тот или иной предложенный образец. Поэтому целесообразно использовать принцип обратной связи пользователя с рекомендующей системой: люди оценивают элементы как интересные или не интересные, а система пытается найти новые элементы, которые подобны интересным и не похожи на не интересные, отрицательные примеры. При достаточном количестве положительных и отрицательных примеров современные методы машинного обучения способны классифицировать новые элементы с высокой

точностью и определить, интересны ли они пользователю.

Рекомендующие системы (РС) отличаются от информационно-поисковых систем (ИПС) тем, что для обнаружения необходимого пользователю результата нет необходимости явным образом формулировать поисковый запрос – система сама, на основании имеющихся сведений о пользователе, предлагает пользователю то, что может его заинтересовать или быть ему полезным. Чем выше и точнее информированность РС о потребностях пользователя, тем более эффективны результаты ее работы.

РС представляют собой один из подходов к решению проблемы информационной перегрузки. Использование технологий Semantic Web и Web 2.0. позволяет интегрировать знания о пользователях и ресурсах, и повысить эффективность формирования рекомендаций, а разработка гибкой стратегии построения рекомендаций, явно формируемой пользователем, обеспечивает более высокую персонализацию РС.

### Типы рекомендующих систем

В [1] РС определяются как программные средства и методы, предлагающие пользователям полезные для них элементы. При этом "элемент" является общим термином, используемым для обозначения того, что система рекомендует пользователям. Чтобы точнее определить РС, следует уточнить ее роль, различая при этом роль РС для поставщика рекомендуемых элементов (РЭ) от ее роли для пользователя.

Поставщики РЭ могут использовать технологии РС для:

- 1) увеличения количества продаваемого товара;
- 2) продажи более разнообразной продукции;
- 3) повышения удовлетворенности пользователей;
- 4) увеличения лояльности пользователей;
- 5) лучшего понимания потребностей пользователей.

С точки зрения пользователей, РС позволяют:

1) найти несколько подходящих пользователю РЭ (ранжированный список и прогноз их полезности);

- найти все РЭ, которые могут удовлетворить некоторые потребности пользователей;
- аннотации найденных РЭ в контексте долгосрочных пользовательские предпочтения;
- рекомендовать последовательность РЭ (порядок чтения книг, просмотра фильмов, выбор обучающих курсов и т.д.);
- рекомендовать комплект – группу РЭ, используемых совместно (например, план отдыха включает транспорт, услуги по размещению и набор экскурсий).

В своей простейшей форме, персонализированные рекомендации – это упорядоченные списки РЭ. При этом упорядочении РС, основываясь на предпочтениях и ограничениях пользователей, пытаются прогнозировать, какие товары или сервисы наиболее подходят пользователю. Для этого РС накапливает информацию о предпочтениях пользователя и о его действиях.

РС широко используются в электронной коммерции, машинном обучении, электронных библиотеках, интегрируются с программными агентами – персональными помощниками и мультиагентными системами. Они применяются на таких широко известных Web-сайтах, как Amazon.com, YouTube, Yahoo и т. д.

Формально формирование рекомендаций в РС может быть представлена следующим образом: пусть  $C$  – множество пользователей РС,  $S$  – множество предлагаемых РЭ (товаров, книг, фильмов, сервисов и т. д.).  $U$  – функция полезности, описывающая интерес пользователя  $c \in C$  к РЭ  $s \in S$ , т. е.  $U : C \times S \rightarrow R$ , где  $R$  – количественная оценка. Цель РС – для каждого потребителя  $c \in C$  выбрать такой РЭ  $s \in S$ , что  $U(c, s) = \max_{s \in S} u(c, s)$ . Каким именно обра-

зом определяется функция полезности, зависит от типа РС и от специфики РЭ.

РС базируются на методах коллаборативной фильтрации, контентной фильтрации и их сочетаниях. При классификации РС обычно выделяют следующие подходы к отбору РЭ:

- *персональный* подход – анализ профиля конкретного пользователя, его ранее проявленных предпочтений и явным образом выраженных условий;

- *социальный* (коллаборативный) подход – анализ предпочтений других пользователей, которые по тем или иным причинам могут распространяться и на того пользователя, для которого делается выбор;

- *контент-ориентированный* подход, при котором анализируются сами РЭ, предлагаемые пользователю;

- *доверительный* подход – анализируется качество предлагаемых пользователю РЭ и анализируется степень доверия к ним.

Следует отметить, что в большинстве реальных РС все эти подходы реализуются интегрировано, но им придается различное внимание.

При контентной фильтрации полезность  $u(c, s)$  определяется на основе значений полезности  $u(c, s_i)$ , определенных этим пользователем для подобных РЭ, а при коллаборативной – на основе значений полезности  $u(c_i, s)$ , определенных для этого РЭ другими пользователями.

Чаще всего контентные РС исполь-

зуют для рекомендования РЭ, содержащих текстовую информацию (например, Web-сайтов, новостей) [2]. Важность слова может определяться, например, с помощью частоты его использования. Значительно сложнее применять контентную фильтрацию к мультимедийным объектам, в которых надо вначале распознавать естественно-языковой текст.

Иногда необходимо отфильтровывают объекты, не только если они несходны с предпочтениями пользователя, но и в том случае, если они обладают слишком большим сходством.

Кроме того, чтобы преодолеть проблему узких рекомендаций, когда пользователь никогда не узнает о тех РЭ, с которыми ранее не сталкивался, можно применить фактор случайности (например, генетические алгоритмы).

Алгоритмы для коллаборативной фильтрации можно разделить на два больших класса: анamnестические (на основе памяти) и модельные.

Анамнестические алгоритмы прогнозируют оценки РЭ клиентом, исходя из всех предшествующих оценок, сделанных данным клиентом, т. е. не оцененный ранее пользователем РЭ получает такую оценку, какую дали ему те пользователи, которые так же (или похоже) оценили ранее оцененные пользователем РЭ.

В большинстве РС подобие между двумя пользователями основывается на том, какие оценки они дали одним и тем же товарам. Наибольшее распространение получили корреляционный метод и метод линейного сходства.

$$U(c^{\wedge}, s^{\wedge}) = \frac{1}{N} \sum_{i=1}^N U(c^{\wedge}, s_i) * \text{sim}(c^{\wedge}, c_i),$$

где  $\text{sim}(c_1, c_2)$  – функция, оценивающая степень подобия пользователей  $c_1, c_2$ ,  $\text{sim}(c_1, c_2) \in \overline{0,1}$ , которая может определяться, например, следующим образом:

$$\text{sim}(c_1, c_2) = \frac{1}{M} \sum_{j=1}^M Q(c_1, c_2, s_j),$$

где

$$Q(c_1, c_2, s) = \begin{cases} 1, & \text{если } U(c_1, s) = U(c_2, s) \\ 0, & \text{если } U(c_1, s) \neq U(c_2, s) \end{cases},$$

а  $M$  – количество РЭ, оцененных обоими пользователями.

Эта функция может быть описана более сложно, если оценки могут иметь не только бинарные значения.

Однако такая функция определения подобия не учитывает того, что пользователи по-разному использовать шкалу оценок, и потому более точный результат дает учет среднего значения оценок, даваемых каждым пользователем.

В отличие от анamnестических методов, модельные алгоритмы прогнозируют оценку РЭ, исходя не из эвристических допущений, а на некоей модели поведения, созданной с помощью статистического анализа, используя, например, кластерные модели, латентно-семантический анализ и Байесовы сети.

Для эффективной работы РС надо предвидеть оценки, исходя из небольшого количества примеров. Для преодоления проблемы разреженности оценок следует при поиске похожих пользователей использовать также сведения из их профилей и обнаруживать пользователей со схожими профилями, например, относящихся к одному демографическому сегменту.

Анализ основных направлений развития современных РС [3] связывает их с использованием онтологий для представления знаний как о пользователях, так и о РЭ. При персональном подходе РС необходимо накопить достаточно сведений о пользователе, чтобы в дальнейшем их обобщать и анализировать. Фоновый мониторинг работы пользователя обеспечивает положительные примеры того, что этот пользователь ищет, не мешая его нормальной работе. Для нахождения отрицательных примеров из наблюдаемого поведения также могут применяться эвристики, (хотя в целом с меньшей точностью). Эта идея лежит в основе тех РС, которые наблюдают за поведением пользователей и рекомендуют им те новые РЭ, которые коррелируют с профилями пользователей. Например, если пользователь регулярно просматривает сайты определен-

ной тематики, то РС, проанализировав контент этих сайтов, может предложить ему другие сайты той же направленности.

Другой способ рекомендовать РЭ базируется на рейтингах, предоставляемых теми людьми, которые ранее оценили РЭ. Коллаборативные РС для этого запрашивают у пользователей явные оценки РЭ, а затем рекомендуют те РЭ, которые высоко оценили похожие пользователи. При коллаборативной фильтрации нет прямого вознаграждения за выставления оценок, которые помогают другим людям.

Профилерование пользователя, т. е. создание формализованного описания пользователя, позволяющего прогнозировать его реакцию на различные РЭ, как правило, базируется либо на знаниях о нем, либо на анализе его поведения. Для получения знаний о пользователе часто используются анкетирование и интервью. После того, как модель для пользователя выбрана, можно использовать знания соответствующей этому типу пользователей предметной области для того, чтобы помочь конкретному пользователю.

Подходы, основанные на поведении, используют поведение пользователя в качестве модели, а обычно используемые методы машинного обучения позволяют обнаружить полезные закономерности в поведении. Регистрация поведения используется для получения данных, необходимых для извлечения закономерностей. В [4] приведен обзор методов создания модели пользователя – его планов, предпочтений, знаний, и методы пополнения этой модели сведениями из новых информационных ресурсов.

В большинстве РС профилерование пользователей базируется на поведении и обычно использует бинарную модель классов для представления того, что пользователь считает интересным и не интересным. Для поиска РЭ, потенциально интересных пользователю, используются методы машинного обучения, рекомендуемые РЭ, которые соответствуют положительным примерам и не соответствуют отрицательным примерам. Но бинарный профиль не позволяет совместно использовать примеры проявления инте-

реса или интегрировать разные области знаний.

Ранние РС, использующие подходы на основе бинарной классификации контента, анализируют обучающую выборку примеров того, что было интересно и не интересно конкретному пользователю. Методы машинного обучения используются для выполнения контролируемого обучения на основе наборов наблюдаемых примеров, которые пользователь обозначил как «хорошо» или «плохо». Классический пример РС на основе контента – система Fab [5], которая использует бинарную классификацию К-ближайшего соседа.

Онтологии позволяют расширить РС для гетерогенной среды, обеспечивая использование основанных на знаниях подходов совместно с классическими алгоритмами машинного обучения, статистическими корреляциями, профилерованием пользователей и специфическими для предметной области эвристиками. Коммерческие РС, как правило, поддерживают простые онтологии продуктов (например, книг), которые они могут использовать при помощи эвристик, или имеют большое сообщество пользователей, активно оценивающих контент (например, фильмы), что делает возможной коллаборативную фильтрацию. РС, ориентированные на исследования, используют значительно больше методов, что обеспечивает повышенную точность, но усложняет работу пользователей (например, навязчивый мониторинг поведения пользователей в течение длительного периода времени).

Рекомендация относительно новых РЭ для пользователей может формироваться на основе его сравнения с подобными РЭ (фильтрация на основе контента), отзывов об РЭ в сообществе пользователей (коллаборативной фильтрации), семантических отношений между РЭ (эвристические рекомендации) или сочетания этих подходов. Во многих случаях выбор подхода зависит от того, насколько доступны метаданные об РЭ и есть ли обратная связь с пользователями (явно и неявно). Методы на основе контента хорошо

работают, если есть достаточная обучающая выборка, а коллаборативные методы – когда система имеет большое сообщество пользователей. Однако на сегодня не выработаны общепринятые правила для выбора стратегии рекомендации: чтобы выбрать наиболее удачный подход для конкретной предметной области как правило, необходимы опыт и знания.

Для расширения бинарной классификации в онтологии домена были введены мультиклассовые классификации и, следовательно, мультиклассовые рекомендации. Обычно классы в онтологии ПрО, такой как онтология товаров на Web-сайте электронной коммерции, используются для классификации товаров, которые пользователь купил ранее, и Web-страниц, которые он просматривал. Примером мультиклассовой рекомендации является Raar [6], который использует простой набор категорий для представления индивидуальных профилей пользователей.

После того, как ПрО классифицирована в терминах онтологических понятий, отношения, определяемые онтологией ПрО, могут использоваться для логического вывода интересности и релевантности одного понятия из наблюдаемого интереса к другому понятию. Основанная на знаниях система может использовать правила экспертной системы для вывода вероятности интереса к классам РЭ, семантической связанным с наблюдаемым интересом к РЭ. Обычно семантическое расстояние (количество отношений от одной теме от другой) используется для расчета семантического сходства, и это используется, чтобы определить вес интереса. Например, в [7] рассматриваются РС, использующие знание-ориентированный подход к формированию рекомендаций о выборе ресторанов.

Если запросы пользователей сформулированы через Web-интерфейс, то критериям запроса может управлять основанное на знаниях дерево решений для уточнений запроса. Примером такого подхода является система CWAdvisor [8], при котором конечное состояние модели используется для уточнения запросов о

доступных финансовых услугах, которые соответствуют установленным требованиям пользователя.

Некоторые области не имеют четко определенных классов РЭ, с помощью которых можно классифицировать контент. В этих случаях рекомендующие системы используют методов кластеризации для выявления в пределах группы РЭ потенциально похожих классов. Например, иерархическая кластеризация использована для классификации коллекций документов [9].

Подразделяют кластеризацию на основе расстояния и на основе понятий.

Кластеризация на основе расстояния использует подход к построению иерархического дерева классов либо сверху вниз (разделение), либо снизу вверх (агломерация). Она также может использовать онтологический подход для учета предпочтений пользователей. Функция расстояния, которая используется для вычисления сходства между документами, часто основывается на подобию частоты использования слов в документе. Алгоритм кластеризации повторяется, подразделяя супер-кластеры или объединяя малые кластеры в более крупные, пока не будет сформировано дерево понятий.

Кластеризация на основе понятий принимает РЭ, представленные в виде пар атрибутов, и строит отношения, основанные на вероятности появления пар атрибутов в узлах. Одним из первых примеров кластеризация на основе понятий является алгоритм COBWEB [10]. Узлы создаются при нисходящем подходе, при котором узлы разделяются или объединяются в соответствии со значением полезности категории; полезность категории является мерой силы дифференцирования этого узла.

Коллаборативная фильтрация использует рейтинги, предоставляемых сообществом пользователей, чтобы рекомендовать РЭ конкретному пользователю. Существуют два взаимодополняющих подхода к коллаборативной фильтрации: на основе пользователя или на основе РЭ. При коллаборативной фильтрации на ос-

нове пользователя находят группы подобных пользователей, а затем конкретному пользователю рекомендуют те РЭ, которые понравились другим пользователям из той же группы. При коллаборативной фильтрации на основе РЭ группируются те РЭ, которые одинаково оцениваются людьми. Для того, чтобы выполнить коллаборативную фильтрацию, должен быть создан профиль пользователя на основании имеющихся документов о том, какие РЭ были этим пользователем рассмотрены и оценены. Обычно для оценки используется 5-балльная шкала (от очень хорошего до очень плохого). Общепринятое представление профиля пользователя представляет собой взвешенный вектор, в котором каждому классу ПрО соответствует его оценка пользователем. Эти векторы могут быть также использованы для определения сходства между РЭ. Для ПрО, в которых метаданные об РЭ не доступны как онтологические термины, как правило, применяют методы предварительной обработки для вычисления частоты терминов в словах, документах и метаданных, удаляют стоп-слова (общие слова всех ПрО) и объединяют похожие слова, используя какой-либо тезаурус (например, WordNet). Коллаборативная фильтрация на основе пользователя – наиболее популярный алгоритм рекомендации благодаря своей простоте и высокому качеству рекомендаций. Вначале отношения соседства формируются с помощью метрики подобия. Затем создается набор прогнозов рейтинга с использованием профилей, которые находятся близко к персональному профилю пользователя. Коллаборативная фильтрация на основе РЭ стала популярной в последние годы, поскольку она разделяет модель вычислений и процесс прогнозирования; в частности, этот метод успешно используется Amazon [11] для обработки сведений о большом количестве как пользователей, так и РЭ. Так же, как и основанном на пользователе методе, сходство РЭ определяется при помощи того, сколько пользователей оценили эти РЭ как подобные. При этом определяются наборы похожих предметов. Этот метод хорошо масштабируется,

поскольку новые РЭ добавляются к окрестностям на основе того, как пользователи оценивают их, без необходимости явного использования онтологии.

Иногда РС должна сравнивать РЭ из разных онтологий ПрО, таких, как два списка товаров. В этих случаях онтологии для обоих доменов могут быть созданы на одном языке (например, OWL) и отображение между ними задается вручную или автоматизировано, например, с помощью байесовской сети доверия: так, в [12] предложен метод на основе ключевых слов, не зависящий от структуры онтологии товаров. Если понятия отображаются успешно, то затем используют обычные подходы для рекомендаций.

Современные исследования также используются некоторые новые ресурсы Web 2.0 и Semantic Web, чтобы помочь в классификации РЭ. Одна из таких систем [13] используется для рекомендации фильмов и работает с базой данных, которая содержит подробную информацию об актерах, фильмах и т.д., и отображает эту семантическую информацию на поведение пользователей на Web-сайте рекомендации фильмов и используя технологии Semantic Web для интеграции сведений из различных источников. Облако тегов создаются на основе частоты ключевых слов, связанных с РЭ, который оценивается. Методы Data mining также могут быть связаны с онтологическими знаниями для улучшения соответствия сходства и рекомендаций с использованием исторических данных.

### Постановка задачи

Анализ существующих подходов к построению рекомендующих систем выявил необходимость в построении формальной классификации методов выработки рекомендаций и обеспечении ее доступности пользователям этих систем. Кроме того, целесообразно разрабатывать методы и стратегии, использующие знания об интересующей пользователя предметной области и позволяющие пользователю непосредственно управлять способом выработки рекомендаций.

## Критерии классификации подходов к выработке рекомендаций в РС

На основе вышеприведенного анализа можно предложить следующие подходы к работе РС с использованием онтологий: формирование модели пользователя; формирование модели РЭ; создание онтологии РЭ; накопление сведений об экземплярах РЭ и экземплярах пользователей; накопление оценок РЭ пользователями; анализ экземпляров РЭ; классификация (или кластеризация) пользователей на группы с подобными интересами; формирование набора стратегий, которые пользователь может явно выбирать для получения рекомендации; построение метода, позволяющего уточнить класс необходимого пользователю РЭ.

Широко распространенные классификации подходов к выработке рекомендаций, подразделяющие все существующие методы на базирующиеся на пользователе и базирующиеся на РЭ, а также на персональные и коллаборативные, являются слишком общими и, как правило, бинарными. Кроме того, в РС большое внимание уделяют алгоритмам вычисления подобия пользователей и РЭ, и значительно меньшее – методам классификации пользователей и РЭ.

На практике целесообразно не только использовать больше критериев, значимых для выработки рекомендаций, но и предоставить пользователю РС возможность самостоятельно формировать стратегию рекомендации, явным образом указывая значимость для данной задачи тех или иных критериев.

При выработке рекомендаций в РС многое зависит от специфики искомого РЭ.

Все РЭ можно подразделить на две категории с точки зрения возможности их *повторного использования*: используемые одноразово и многократно. К первой категории относятся различные предметы материального мира и связанные с их использованием услуги. К ним относятся, например, технические устройства, продукты питания, авиабилеты, турпоездки. Если у пользователя уже есть такой РЭ, то

ему может понадобиться такой же или похожий на него (после поломки первого, использования и т.д.). К второй категории относятся, как правило, информационные объекты, т.е. такие объекты, что наличие одного экземпляра позволяет создавать произвольное количество его копий. К ним относятся, например, электронные книги и фильмы. Если у пользователя уже есть такой РЭ, то маловероятно, что ему понадобится еще один (хотя возможна утрата или поломка).

Кроме того, РЭ можно классифицировать (на два или более классов – в зависимости от требуемой точности рекомендаций) на *редко или часто* используемые. К примеру, прогноз погоды нужен пользователю почти ежедневно, а выбор модели холодильника актуален для большинства раз в 10 лет. Для наиболее часто используемых РЭ большинство пользователей склонны ориентироваться на собственный опыт, а для редко используемых – на совокупный опыт сообщества пользователей РС. При этом следует учитывать, что один и тот же РЭ может одним пользователям быть интересен редко, а другим – часто. Например, большинство людей редко интересуется особенностями газовых плит или мебельной фурнитурой, но специалист по комплексным ремонтам может выполнять такие запросы регулярно. Поэтому надо дать пользователю возможность самому явно определять, насколько часто он интересуется такими РЭ (т.е. насколько его собственное мнение о них компетентно и актуально).

Еще один важный параметр РЭ – *субъективность* оценивания. Если, к примеру, при оценке бытовой техники или автомобилей достаточно легко сформулировать те отличия, по которым пользователь более высоко оценивает один РЭ, чем другой (например, надежность работы, стоимость, простота обслуживания, функциональные возможности), и вследствие этого каждому пользователю может быть полезен опыт всего сообщества, то при оценке предметов искусства, музыки, фильмов такие отличия практически невозможно формализовать, и потому при выработке рекомендаций для конкретного пользова-

теля важен только опыт некоторого подмножества сообщества с аналогичными вкусами, причем это подмножество может быть сформировано с помощью методов машинного обучения и на основе продуктивного вывода.

На способ выработки рекомендаций влияет и то, насколько оценивание РЭ требует *специальных знаний* в конкретной предметной области. Особенно важно это для тех РЭ, которыми большинство пользователей интересуются редко и потому сами, как правило, не имеют о них глубоких знаний. Вследствие этого они оценивают, как правило, лишь конкретный предмет (а не все предметы данного класса) и практически не имеют возможности сравнить его с другими подобными РЭ. Например, оценивая телевизор, пользователь может оценить лишь ту модель телевизора, которую он купил, и сравнить ее лишь с несколькими теми моделями, которыми он пользовался. Поэтому в некоторых областях важнее ориентироваться на мнение экспертов, а не на мнение большинства.

Так, при выборе учебника по какой-то дисциплине больший вес имеет оценка преподавателя, при покупке бытовой техники – мнение специалиста по ремонту, а при подборе лекарства – оценка врача. Открытым остается вопрос о том, каким образом формировать множество экспертов для той или иной ПрО. В частности, в большинстве социальных сетей существует как возможность зафиксировать как связь пользователя с набором тем, так и средства определения общего рейтинга пользователя (влиятельность его мнения для других пользователей, оценка его действий другими пользователями и т.д.), но, как правило, отсутствует возможность дифференцировать компетентность пользователя в той или иной тематике.

Например, один и тот же пользователь, проявляющий интерес к компьютерам и кулинарии, может быть экспертом в компьютерной технике (и его оценки различных экземпляров компьютеров будут высоко точными), а в приготовлении пищи разбираться очень плохо (и именно потому и интересоваться этой областью) и потому

оценивать различные РЭ крайне неправильно.

Следует отметить, что в процессе развития Web и систем электронной коммерции появилось много источников, обеспечивающих доступ к одним и тем же РЭ. При этом речь идет как о материальных, так и об информационных продуктах и услугах. Но при этом разные источники предлагают разные условия доступа и качества обслуживания. Поэтому значительного внимания заслуживает степень доверия к источникам, которую можно определить по совокупности оценок сообщества пользователей. Этот критерий рекомендации выделяется отдельно, так как речь идет не об оценивании РЭ, а об оценивании *источников РЭ*. Например, многие Web-сайты предлагают бесплатно скачивать статьи, книги и фильмы, но некоторые из них требуют регистрации, отправки SMS-сообщений (не бесплатных), оплаты пароля на распаковку архива и т.д. очевидно, что такие сайты оцениваются значительно ниже, чем те, которые предлагают свободный доступ к электронной библиотеке. Электронные магазины, предлагающие покупку материальных предметов, также работают не всегда честно – они могут задерживать доставку, предоставлять бракованную продукцию, требовать дополнительной оплаты доставки, значительно завышать цену по сравнению с обозначенной (например, на сайте цена представлена в долларах, а оплату следует осуществлять в гривнах по крайне невыгодному курсу, отличающемуся от официального). Поэтому следует предоставить пользователям оценить такие магазины ниже, чем те, которые работают корректно.

В большинстве используемых на практике РС вырабатываются рекомендации относительно какого-то довольно узкого класса РЭ, и нет возможности связывать профили одного и того же пользователя, сформированные различными РС (например, нельзя связать предпочтения пользователя относительно выбора художественной литературы, покупок в электронном магазине и при просмотре новостей). Использование онтологических моделей пользователей позволяет в какой-то

мере решить эту задачу и интегрировать различные РС. При этом возникает дополнительная задача – классификация РЭ, относительно которого пользователь нуждается в рекомендации (и, соответственно, переадресация запроса к соответствующей специализированной РС). Следует учитывать, что нередко пользователь нуждается в рекомендациях по набору взаимосвязанных вопросов (например, выбор места для отдыха связан и с выбором турпутевки, и с прогнозом погоды, и с рекомендациями относительно транспорта). Для этого необходимо разработать общую онтологию РЭ, которая должна быть достаточно компактной и несложной, но при этом охватывать основные классы РЭ, относительно которых пользователи часто нуждаются в рекомендациях. Специфические знания предметных областей такая онтология не должна включать, т. к. они должны содержаться в онтологиях специализированных РС.

Рассмотрим также, оценки каких именно групп пользователей целесообразно применять для коллаборативной фильтрации. Самый простой случай группы – это группа, состоящая всего из *одного пользователя*, для которого и осуществляется поиск рекомендаций. Можно сказать, что при этом коллаборативная фильтрация сводится к персональной. Но и в этом случае можно ввести ряд управляемых пользователем опций – например, учитывать его опыт только за определенный период времени (например, пользователь считает собственные оценки компьютерной техники, сделанные более 5 лет назад, уже не актуальными). Это особенно важно для быстро изменяющихся предметных областей и для пользователей младшего возраста (очевидно, что в 10 и 15 лет интересы и оценки пользователя могут быть различными).

Противоположный случай – когда для пользователя значимы оценки РЭ *всем сообществом* в целом. Это может иметь место для тех ПрО, к которым пользователь обращается впервые и еще не имеет собственного мнения не только о самой области, но и о критериях нахождения в ней экспертов.

К промежуточным случаям относятся анализ оценок экспертов в ПрО (причем тоже надо предоставить пользователю возможность явно задавать приемлемый уровень их квалификации).

В целом следует оценивать выбранную стратегию рекомендации по трем направлениям – учет мнения самого пользователя, учет мнения сообщества, анализ самого РЭ. В таком трехмерном пространстве (рисунок) можно разместить большинство типичных объектов рекомендации.

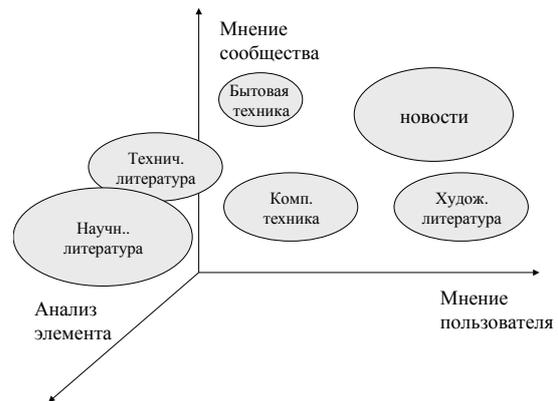


Рисунок. Классификация стратегий рекомендации

Выбор пользователем значения по каждому из трех параметров для стратегии рекомендации для интересующих его РЭ и наличие онтологии (или хотя бы таксономии РЭ) позволяет достаточно точно профилировать интересы самого пользователя, оценить его собственную компетентность для оценивания РЭ (и, соответственно, значимость его мнения для других пользователей) и выявлять группы пользователей со сходными интересами. Следует отметить, что более объективным является признание своей некомпетентности в оценивании РЭ, чем декларирование своей высокой квалификации.

Существует несколько направлений усовершенствования работы РС:

- 1) расширение профилей пользователей и РЭ (на основе Text Mining, анализа сетевого поведения и т. д.);
- 2) многомерность рекомендаций (вместо традиционного двумерного пространства «Пользователь-РЭ»), учитыва-

ющих контекст – например, время года или суток, погоду, местонахождение пользователя;

3) ненавязчивость – возможность уточнить рекомендации по косвенным предпочтениям.

### Тезаурусы как средство представления знаний о задаче

Следует отметить, что важным требованием к РС является понятность для пользователя рассуждений системы. В общем случае онтология ПрО является достаточно сложной структурой, и потому при решении конкретной задачи целесообразнее использовать только часть знаний, содержащихся в такой онтологии и представленных в более простой для понимания форме – тезаурус задачи, который можно рассматривать как проекцию онтологии на задачу. Именно такой тезаурус затем используется для анализа контента текстовых РЭ и метаописаний всех остальных РЭ РС.

Формирование онтологии ПрО без дополнительных ограничений – достаточно сложный, трудоемкий и длительный процесс. Поэтому представляется целесообразным использовать для моделирования знаний пользователя об интересующей его ПрО частного случая онтологии – тезауруса, построение которого относительно проще. Тезаурус из всего спектра средств языка отражает только лексику: она задана в знаковом виде и относительно просто поддается систематизации. Тезаурус можно представить как комплекс лингвистических знаний, включающий все составляющие языка от фонетики до риторической структуры текста и законов коммуникации [14].

Обычно тезаурус  $T$  определяют как словарь, содержащий лексические единицы с явным указанием семантических связей между ними. Еще в начале 60-х гг. прошлого века Ю.А. Шрейдер предлагал рассматривать тезаурус как систему знаний, отраженных языком, когда тезаурус становится интересным сам по себе, а не только как вспомогательный инструмент.

Тезаурус в ИТ – это полный систематизированный набор данных о какой-

либо области знаний, позволяющий человеку или вычислительной машине в ней ориентироваться. Тезаурус – это  $Ts = \langle T, R \rangle$ , где  $T$  – множество терминов, а  $R$  – множество отношений между этими терминами. Множества  $T$  и  $R$  конечны.

Можно рассматривать тезаурус как модель терминологической системы. Терминологическая система (ТС) – это сложная динамическая устойчивая система, РЭ которой являются отобранные по определенным правилам лексические единицы какого-нибудь естественного языка, а структура изоморфна структуре логических связей между понятиями специальной области знаний и деятельности, а функция состоит в том, чтобы служить знаковой (языковой) моделью этой области знаний и деятельности [15]. Можно говорить о том, что ТС – отображение определенной ПрО.

Множество терминов тезауруса  $T$  соответствует множеству концептов  $X$  онтологии  $O$ . Такие свойства терминов и ТС, как системность, устойчивость и регулярность связей, отсутствие экспрессии, установка на объективность описания, делают возможным моделирование ТС с помощью тезаурусов. Классификация понятий ПрО через набор слов, условно синонимичных и образующих класс условной эквивалентности, лежит в основе тезаурусов, используемых для информационного поиска.

Многие авторы рассматривают тезаурус как иерархическую структуру или ряд иерархических структур, в которые сводятся группы понятий с общими характеристиками, выраженные в терминах естественного языка. Для большинства реальных ПрО это не так, поскольку различные отношения, используемые в ПрО, несравнимы (например, "статья написана автором", "программа разработана на языке"). Более корректно тезаурус можно представить как семантическую сеть. В узлах такой сети находятся термины, связанные отношениями из ограниченного набора. Для представления тезауруса целесообразно также использовать ориентированные графы и фреймы. Сохраняя многие из достоинств семантической сети, фреймовая сеть позволяет в качестве вершин

представить сложные структуры (фреймы), имеющие, в частности, незаполненные поля (слоты), что дает новые возможности при описании вложенных структур, переключении между различными приложениями и т. д.

Большинство существующих информационно-поисковых систем имеют развитые средства контекстного поиска документов с учетом морфологической информации о словах. Однако в настоящее время очень незначительное число информационных систем предоставляют возможность тематического поиска, например, поиска с использованием тезауруса. Каждое понятие в тезаурусе может объясняться через набор других понятий, что приводит к появлению семантического поля. Фактически тезаурус пользователя-потребителя информации – это вербализованная совокупность его представлений об исследуемой ПрО. Основной целью разработки информационно-поисковых тезаурусов является использование их единиц (дескрипторов) для описания основных тем документов в процессе ручного индексирования.

Тезаурусы позволяют моделировать знания как о пользователях РС, так и о РЭ [16].

Чтобы уведомить РС об области своих интересов – ПрО поиска – пользователь должен создать тезаурус, моделирующий интересующую его ПрО, в котором содержатся основные термины ПрО и связи между ними. Тезаурус можно создать вручную или автоматизированно. Основой для автоматического создания тезауруса может послужить обработка набора текстовых документов – информационных ресурсов (ИР), которые пользователь считает релевантными этой ПрО, или ранее созданная онтология ПрО, из которой пользователь отбирает только необходимые ему термины. Все эти подходы могут комбинироваться друг с другом.

Для необходимости анализа большого количества ИР предлагается использовать упрощенный алгоритм построения тезауруса: по полному перечню слов, используемых в ИР, строится словарь терминов, из которого отбрасываются стоп-

слова, содержащиеся в специально разработанном пользователем списке. Этот алгоритм применяется только для тех ИР, которые не сопровождаются метаописаниями. В противном случае из метаописаний (в формате RDF или OWL) извлекаются термины тезауруса и связи между ними, которые дополняют построенный по контенту ИР словарь.

Аналогично строятся тезаурусы РЭ – обрабатываются их метаописание, контент, отзывы о них других пользователей.

Для выбора тех РЭ, которые будут рекомендоваться пользователю, необходимо сопоставить их тезаурусы с тезаурусами пользователей.

Пользователь вводит запрос, приблизительно идентифицируя свою информационную потребность с помощью ключевых слов или выбирая класс интересующего его РЭ (возможно, с набором условий и ограничений), например, РЭ класса «художественная литература/фантастика/фэнтези», изданная после 2005 года. В ответ РС формирует набор РЭ, доступных системе и соответствующих этому приблизительно запросу –  $n$  ссылок на РЭ и их кратких описаний  $I = \{Re f_j, D_j\}$ ,  $j = \overline{1, n}$ . Здесь  $Re f_j$  – ссылка на соответствующий РЭ (или его описание), а  $d_j$  – информация об этом РЭ,  $t_{j_lw}$  доступная РС.

Если множество  $I$  не пусто, причем РС найден в ответ на запрос более чем один РЭ ( $n \geq 1$ ), то нужно установить порядок, в каком предлагать пользователю сведения о найденных РЭ. Тогда для всех РЭ из этого множества  $I = \{Re f_j, D_j\}$ ,  $j = \overline{1, n}$  формируются их упрощенные тезаурусы  $Ts(ИР_j) = \langle T_{j_j}, \emptyset \rangle$ ,  $j = \overline{1, n}$  и соответствующие им словари терминов  $T_j = \{t_{j_lw}\}$ ,  $j = \overline{1, n}$ ,  $w = \overline{1, q_j}$ .  $t_{j_lw}$  – это слова, которые используются в информации о  $j$ -м РЭ, найденном РС, т. е. в  $D_j$ ,  $j = \overline{1, n}$ .  $q_j$ ,  $j = \overline{1, n}$  – это количество различных слов, используемых в описании  $D_j$ ,  $j = \overline{1, n}$ . Если слова в описании повто-

ряются, то в словаре терминов они фиксируются только один раз.

Затем пользователь формирует тезаурус интересующей его ПрО (или указывает на ранее сформированный тезаурус)  $T_{s_{ПрО}}$  и соответствующий ему словарь терминов этой ПрО  $T_{ПрО} = \{t_m\}$ ,  $m = \overline{1, q}$ .  $T_{ПрО}$  – это множество, состоящее из  $m$  терминов, относящихся к интересующей пользователя ПрО. Это множество строится аналогично словарю терминов РЭ и обычно формируется как объединение словарей терминов, содержащихся в документах, которые пользователь нашел ранее и посчитал релевантными интересующей его ПрО (как в их контенте, так и в метаописаниях).

Производится сравнение  $T_{ПрО}$  и  $T_j$ ,  $j = \overline{1, n}$ , высчитывается коэффициент их близости

$$K_j = \sum_{m=1}^q \sum_{w=1}^{w_j} f(t_{j_w}, t_m), \quad m = \overline{1, q}, \quad w = \overline{1, w_j},$$

где

$$f(t_1, t_2) = \begin{cases} 0, & \text{если } t_1 \neq t_2, \\ 1, & \text{если } t_1 = t_2. \end{cases} \quad (1)$$

Коэффициент (1) представляет собой количество терминов, которые встретились как в тезаурусе РЭ, так и в тезаурусе ПрО. Найденные ИР упорядочиваются в зависимости от значений  $K_j$ , пользователю предъявляются в первую очередь те ИР, которые имеют наиболее высокий коэффициент близости к ПрО.

При использовании коэффициента (1) возникает следующая проблема: слова, соответствующие одному термину, но являющиеся, например, различными словоформами, синонимами или переводами на различные языки, обрабатываются как разные термины. Поэтому представляется целесообразным использовать онтологию ПрО и выделять группы слов, соответствующих одному термину. Для этого пользователь должен связать РЭ словаря терминов тезауруса ПрО с одним из терминов онтологии ПрО  $O = \langle X, R, F \rangle$ , т. е.

$\forall t_m \in T_{ПрО}, \quad m = \overline{1, q}$  задать функцию  $g(t_m) \in X$ . Затем для вычисления коэффициента близости  $K^O$  эта функция используется следующим образом:

$$K^O_j = \sum_{m,l} f(t_{j_w}, t_m), \quad m = \overline{1, q}, \quad w = \overline{1, w_j},$$

где

$$f(t_1, t_2) = \begin{cases} 0, & \text{если } g(t_1) \neq g(t_2), \\ 1, & \text{если } g(t_1) = g(t_2). \end{cases} \quad (2)$$

Коэффициент (2) представляет собой количество терминов, которые встретились как в тезаурусе РЭ, так и в тезаурусе ПрО и при этом ссылаются на один и тот же термин онтологии ПрО. По сравнению с коэффициентом (1) коэффициент (2) позволяет использовать меньший объем документов для построения тезауруса ПрО, но требует большее время для вычислений.

При создании тезауруса ПрО, которая интересует пользователя РС, необходимо явно указать основные понятия ПрО и связи между ними. К сожалению, большинству пользователей достаточно сложно это сделать (даже имея соответствующие знания и применяя их в своей деятельности). На первом этапе формирования тезауруса пользователь может выбрать одно из следующих решений: 1) самостоятельно построить с помощью одного из редакторов онтологий онтологическое описание области его информационных интересов; 2) найти (например, в Интернете) какую-либо онтологию, представленную на языке OWL, которую описывает ПрО, близкую к области его информационных интересов; 3) сформировать множество понятий ПрО, которое содержит наиболее характерные слова и словосочетания, встречающиеся в интересующих его ИР.

Важно определить, какие именно связи между РЭ ПрО являются существенными (и их, следовательно, необходимо включить в систему). Не все существенные связи между терминами ПрО могут быть очевидны пользователю, поэтому для их выявления он может воспользоваться для

их нахождения *методами индуктивного вывода*.

Существуют независимые подходы к реализации подобных методов: ID3, ACLS, CART и т. д. Наиболее интересным, в связи со спецификой проводимой работы, оказался алгоритм ID3 [17], который специально разработан для извлечения ценной информации из больших объемов слабо структурированных данных. При работе этого алгоритма время вычислений зависит линейно от числа введенных примеров, числа атрибутов, используемых для описания примеров, и числа узлов в строящемся дереве решений. Это качество отличает его от таких известных алгоритмов построения деревьев решений, как INDUCE, SPROUTER, ROTH-P, в которых усилия, требующиеся для решения задачи, резко возрастают вместе со сложностью задачи.

Если методы, подобные МГУА (метод группового учета РЭ), предназначены для нахождения закономерностей по набору количественных измерений параметров и полученному по ним результату, то методы, подобные ID3 и его вариациям (C4.5, ID4 и т.д.), предназначены для обобщения опыта экспериментов, параметры и результаты которых описаны через качественные оценки (лингвистические переменные). В большинстве случаев между их значениями невозможно установить даже относительное упорядочение (например, различные симптомы и диагнозы пациентов). К таким задачам относятся и проблема, которую решают рекомендуемые системы. ID3 принадлежит к невозрастающим алгоритмам, то есть при добавлении к набору классифицированных примеров определенного количества новых нужно обрабатывать снова как старые, так и новые примеры. Но ID3 предназначен для построения только бинарного дерева решений, а этого недостаточно удобно для представления закономерностей многих ПрО.

Поэтому предлагается использовать *ID3m* [18] – модификацию ID3 для произвольного (конечного) количества решений. Он также принадлежит к невозрастающим

алгоритмам. В данном случае, примерами обучающей выборки являются РЭ, доступные РС, а параметрами, по которым они описываются, являются их свойства, описанные в метаданных и в онтологии РЭ, а также термины тезауруса пользователя. Значения, соответствующие терминам тезауруса, – "Термин отсутствует в описании РЭ", "Термин встречается в описании РЭ редко", "Термин встречается в описании РЭ часто". В качестве результата используется оценка, данная пользователем найденному РЭ (качественная оценка, имеющая два и более значений).

На вход алгоритма поступает обучающая выборка  $H$  – набор из  $n$  классифицированных (получивших одну из возможных оценок) примеров одинаковой размерности  $H = \{h_i\}, i = \overline{1, n}$ .

Каждый пример из выборки представляет собой упорядоченную последовательность значений  $s$  атрибутов и результирующего атрибута  $h_i = \langle a_1, \dots, a_s, r \rangle, i = \overline{1, n}$ . Значения атрибутов принадлежат конечным множествам:  $a_{ju} \in A_j, j = \overline{1, n}, u = \overline{1, n_j}, r_y \in R, y = \overline{1, n_r}$ .

Если обучающая выборка содержит примеры, в которых все значения атрибутов одинаковы, а решения различны, то введенная информация недостаточна для построения классификационного правила. Если множество примеров пустое, то можно произвольно связать его с любым решением. Если все примеры относятся к одному классу, строится один лист дерева решений, связанный с этим классом. В противном случае необходимо выбрать один из атрибутов и разделить множество атрибутов на подмножества в зависимости от значения этого атрибута и применить алгоритм к каждому из полученных подмножеств.

На каждом шаге работы алгоритма вычисляется, какой атрибут  $m$  несет наибольшее количество информации о результате.

$$C_{\max} = \max \{C_z, z = \overline{1, s}\}$$

$$= \max_z \left\{ \sum_i \sum_j \frac{C(a_{z_i} \in A_z, r_j \in R_j)}{d_z} \right\}, \quad (3)$$

где  $C(x, y)$  – количество информации  $C(x, y) = \sum_i \sum_j p(x, y) * \lg p(x, y)$ ,  $p(x, y)$  – вероятность одновременного наступления событий  $x$  и  $y$ ,  $d_m$  – стоимость получения значения  $m$ -го атрибута.

В результате работы алгоритма ID3m формируется дерево решений, в котором каждый лист связан с одним из решений, каждый узел характеризуется именем одного из атрибутов, а выходящие из такого узла ветви – значениями этого атрибута.

Такое дерево решений позволяет РС по параметрам вновь найденного РЭ прогнозировать, как именно оценит его пользователь, и предлагать пользователю в первую очередь те РЭ, которые соответствуют его индивидуальным предпочтениям. Так как точные значения вероятностей событий из обучающей выборки неизвестны, то они аппроксимируются на основе рассматриваемого множества примеров.

Предложенный выше подход к формированию рекомендаций основывается на использовании знаний пользователя о ПрО, характеризующей его информационные потребности. Пользователь может явно указывать интересующие его термины и получать те РЭ, которые соответствуют его потребности. Такой подход ориентирован на пользователя с относительно стабильными информационными потребностями, не являющегося специалистом в области информационных технологий, и позволяет пользователю избежать рутинной работы по фильтрации результатов поиска в Web.

### Методы выработки рекомендаций в МАИПС

МАИПС – мультиагентная информационно-поисковая система с развитыми средствами интеллектуализации ее поведения, которая более детально описана в [19], предназначена для поиска информации в описанных пользователем относительно узких предметных областях,

связанных с профессиональными или научными интересами пользователей, и рекомендует пользователю те результаты поиска, которые относятся к интересующей его предметной области и соответствуют его информационным потребностям. Ее можно рассматривать как рекомендующую систему, ориентированную на формирование рекомендаций относительно естественно-языковых и мультимедийных информационных ресурсов, доступных через Web.

Система МАИПС ориентирована на пользователей, имеющих в сети постоянные информационные интересы и требующих постоянного поступления соответствующей информации. Для этого МАИПС позволяет сохранять и повторно выполнять запросы, учитывая реакцию пользователя на ранее предложенные ему ИР (персональная фильтрация), отслеживать появление аналогичных запросов у других пользователей (коллокативная фильтрация), сохранять формальное описание области интересов пользователя в виде онтологии (семантическая фильтрация) и т.д. Кроме того, в МАИПС при профилировании пользователей используется специфичный для естественно-языковых ИР критерий оценивания – сложность текста для понимания. Особенностью системы является использование оригинального знание-ориентированного алгоритма, позволяющего определить сложность понимания текста для конкретного пользователя (для этого используются тезаурусы предметных областей, интересующих пользователей) [20].

Основой МАИПС являются технологии Semantic Web, в частности, язык представления онтологий OWL и средства его обработки. Для представления знаний об интересующей пользователя ПрО используются онтологии и тезаурусы ПрО. При этом тезаурус строится пользователем по соответствующей онтологии самостоятельно, а онтология выбирается из набора предложенных на сайте.

По мере развития МАИПС возникла потребность в подключении репозитория онтологий, чтобы пользователи могли повторно использовать знания о ПрО, до-

ступные в Web [21]. При этом поиск может осуществляться не только по ключевым словам, а и по другим важным свойствам онтологий. Поэтому в дальнейшем представляется целесообразным реализовать в МАИПС средства взаимодействия с репозиториями онтологий, поддерживающие поиск нужной пользователю онтологии, обнаружение похожих на выбранную пользователем онтологий, а также сопоставление построенного пользователем тезауруса с другими онтологиями и тезаурусами.

Пользователь МАИПС может обращаться к онтологиям, созданных другими пользователями – пересматривать их, задавать по ним контекст поиска, копировать из них нужные фрагменты, но не имеет права изменять их. ИПС может обеспечить поиск онтологий, которые содержат введенные пользователем термины, а также поиск онтологий, похожих на выбранную пользователем онтологию. Это позволяет создавать группы пользователей с общими информационными интересами и предотвратить дублированию в выполнении одинаковых многообразных запросов разных пользователей.

### Выводы

Объем ресурсов, предлагаемых современным Web, требует использования семантических методов для поиска той информации, которая необходима конкретному пользователю. Совместное использование рекомендующих систем и технологий Semantic Web позволяет обеспечить пользователя необходимыми сведениями, а явным образом выбранные методы рекомендации объясняют пользователю поведение такой системы.

1. Ricci F., Rokach L., Shapira B., Kantor P. Recommender Systems Handbook. – Springer, 2011. – 842 p.
2. Adomavicius G., Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions // IEEE Transactions on Knowledge and Data Engineering. – 2005. – Vol. 17, N 6. – <http://artpragmatica.ru/rs/in/pic/58-870-20061024072441>.
3. Middleton S., De Roure D., Shadbolt N. Ontology-Based Recommender Systems // in Handbook on Ontologies, Edt. by S.Staab, R.Studer, Springer, 2009. – P. 779–796.
4. Kobsa A. User modeling: recent work, prospects and hazards. – <http://zeus.gmd.de/~kobsa/papers/1993-aui-kobsa.pdf>.
5. Balabanovic M., Shoham Y. Fab: Content-based, collaborative recommendation // Communications of the ACM, 1997. – 40(3). – P. 67–72.
6. Eirinaki M., Lampos C., Paulakis S., Vazirgiannis M. Web personalization integrating content semantics and navigational patterns // Proc. of the 6th annual ACM International workshop on Web information and data management, USA. – 2004. – P. 72–79.
7. Burke R. Knowledge-based Recommender Systems // In Kent A (ed.) Encyclopedia of Library and Information Systems, V. 69, supplement 32. Marcel Dekker, New York, 2000. – <http://josquin.cti.depaul.edu/~rburke/pubs/burke-elis00.pdf>.
8. Felfernig A., Friedrich G., Jannach D., Zanker M. An integrated environment for the development of knowledge-based recommender applications // International Journal of Electronic Commerce. – 2006. – 11(2). – P. 11–34.
9. Schickel-Zuber V., Faltings B. Inferring User's Preferences using Ontologies. – <http://infoscience.epfl.ch/record/85768/files/AAI06-222.pdf>.
10. Fisher D. Knowledge acquisition via incremental concept clustering // Machine Learning 2(2): 1987. – P. 139–172. – <http://link.springer.com/article/10.1007%2FBF00114265#page-1>.
11. Linden G., Smith B., York J. () Amazon.com recommendations: Item-to-Item collaborative filtering // IEEE Internet Computing. – 2003. N 7(1). – P. 76–80. – <http://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf>
12. Lee T., Chun J., Shim J., Lee S. An ontology-based product recommender system for B2B marketplaces // International Journal of Electronic Commerce. – 2006. – N 11(2). – P. 125–155.
13. Szomszor M., Cattuto C., Alani H., O'Hara K., Baldassarri A., Loreto V. Servedio VDP Folksonomies, the Semantic Web, and Movie

- Recommendatio // Proc. of 4th European Semantic Web Conference, Bridging the Gap between Semantic Web and Web 2.0, 2007. – <http://eprints.soton.ac.uk/264007/1/ESWC2007.pdf>.
14. *Гладун А.Я., Рогущина Ю.В.* Основы методологии формування тезаурусів з використанням онтологічного та мереологічного аналізу // Искусственный интеллект. – 2008. – № 5. – С. 112–124.
  15. *Браславский П.И., Гольдштейн С.Л., Ткаченко Т.Я.* Тезаурус как средство описания систем знаний // Информационные процессы и системы. – 1997. – № 11, Серия 2. – С. 16–22.
  16. *Гладун А.Я., Рогущина Ю.В.* Онтологии и мультилингвистические тезаурусы как основа семантического поиска информационных ресурсов Интернет // The Proc. of XII-th Intern. Conf. KDS'2006, Varna, Bulgaria. – P. 115–121.
  17. *Quinlan J.R.* Discovery rules from large collections of examples: a case study // Expert Systems in the Microelectronic Age. – Edinburgh, 1979. – P. 87–102.
  18. *Рогущина Ю.В.* Применение методов индуктивного вывода для создания прикладных экспертных систем // Разработка и использование информационных технологий в системах управления. – Киев: Ин-т кибернетики имени В.М. Глушкова НАН Украины. – 1993. – С. 122–128.
  19. *Рогущина Ю.В., Гришанова І.Ю.* Літературний твір наукового характеру "Модель мультиагентної інформаційно-пошукової системи "МАІПС" ("Модель МАІПС"). – Свідоцтво про реєстрацію авторського права на твір № 32068, 2010.
  20. *Рогущина Ю.В.* Использование критериев оценки удобочитаемости текста для поиска информации, соответствующей реальным потребностям пользователя // Проблеми програмування. – 2007. – № 3. – С. 76–87.
  21. *Рогущина Ю.В.* Использование репозитивов онтологий для отображения области интересов пользователя при семантическом поиске // XI Междунар. конф. им. Т.А. Таран "Интеллектуальный анализ информации ИАИ-2012". – 2012. – С. 134–140.

**Об авторе:**

*Рогущина Юлия Витальевна,*  
кандидат физико-математических наук,  
старший научный сотрудник.

**Место работы автора:**

Институт программных систем  
НАН Украины,  
03187, Киев-187,  
проспект Академика Глушкова, 40.  
E-mail: [ladamandraka2010@gmail.com](mailto:ladamandraka2010@gmail.com)

Получено 10.10.2012